

On Bounds for E -capacity of DMC *

Evgueni A. Haroutunian, *Associate Member, IEEE*

Abstract—Random coding, expurgated and sphere packing bounds are derived by method of types and method of graph decomposition for E -capacity of discrete memoryless channel (DMC). Three decoding rules are considered, the random coding bound is attainable by each of the three rules, but the expurgated bound is achievable only by maximum-likelihood decoding. Sphere packing bound is obtained by very simple combinatorial reasonings of the method of types. The paper joins and reviews the results of previous hard achievable publications.

Index Terms—Capacity, E -capacity, error probability bounds, rate-reliability function, method of types, method of graph decomposition, decoding rules.

I. INTRODUCTION

Let \mathcal{X}, \mathcal{Y} be finite sets and $W = \{W(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ be a stochastic matrix.

Definition 1: A discrete channel W with input alphabet \mathcal{X} and output alphabet \mathcal{Y} is defined by stochastic matrix of transition probabilities $W : \mathcal{X} \rightarrow \mathcal{Y}$. An element $W(y|x)$ of the matrix is a conditional probability of receiving the symbol $y \in \mathcal{Y}$ on the channel's output if the symbol $x \in \mathcal{X}$ is transmitted from the input.

The model for N actions of the channel W is described by the stochastic matrix

$$W^N : \mathcal{X}^N \rightarrow \mathcal{Y}^N,$$

the element of which $W^N(\mathbf{y}|\mathbf{x})$ is a conditional probability of receiving vector $\mathbf{y} \in \mathcal{Y}^N$, when $\mathbf{x} \in \mathcal{X}^N$ is transmitted. Here we consider memoryless channels, which operate at each moment of time independently of the previous or next transmitted or received symbols, so for all $\mathbf{x} \in \mathcal{X}^N$ and $\mathbf{y} \in \mathcal{Y}^N$

$$W^N(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^N W(y_n|x_n).$$

Let \mathcal{M} denotes the set of messages to be transmitted and M – the number of messages.

Definition 2: A code (f, g) for the channel W is a pair of mappings, where $f : \mathcal{M} \rightarrow \mathcal{X}^N$ is encoding and $g : \mathcal{Y}^N \rightarrow \mathcal{M}$ is decoding. N is called the code length, and M is called the code volume.

* The author is with the Institute for Informatics and Automation Problems of the Armenian National Academy of Sciences, 1 P. Sevak str., Yerevan 0014, Armenia. E-mail: evhar@ipia.sci.am.

Definition 3: The probability of erroneous transmission of the message $m \in \mathcal{M}$ by the channel using code (f, g) is defined as

$$\begin{aligned} e(m, f, g, N, W) &\stackrel{\Delta}{=} W^N (\mathcal{Y}^N - g^{-1}(m)|f(m)) = \\ &= 1 - W^N (g^{-1}(m)|f(m)). \end{aligned} \quad (1)$$

We shall consider the maximal probability of error of the code (f, g) :

$$e(f, g, N, W) \stackrel{\Delta}{=} \max_{m \in \mathcal{M}} e(m, f, g, N, W),$$

and the optimal maximal probability of error for the channel W :

$$e(M, N, W) \stackrel{\Delta}{=} \min_{(f,g)} e(f, g, N, W),$$

where minimum is taken among codes (f, g) of volume M , and the average probability of error for equiprobable messages is

$$\bar{e}(f, g, N, W) \stackrel{\Delta}{=} \frac{1}{M} \sum_{m \in \mathcal{M}} e(m),$$

with $\bar{e}(M, N, W)$ as the minimal average probability among all possible codes of the length N and the volume M . It is clear that always

$$\bar{e}(f, g, N, W) \leq e(f, g, N, W).$$

Definition 4: The transmission rate of a code (f, g) of volume M is

$$R(f, g, N) \stackrel{\Delta}{=} \frac{1}{N} \log M. \quad (2)$$

Note that in this paper all exp-s and log-s are to the base two.

We consider the codes, error probability of which exponentially decrease with exponent E , when $N \rightarrow \infty$:

$$e(f, g, N, W) = \exp\{-NE\}. \quad (3)$$

Denote the best volume of the code of length N for channel W satisfying the condition (3) for given reliability $E > 0$ by $M(E, N, W)$.

Definition 5: The rate-reliability function, which by analogy with the capacity we call **E -capacity**, is for maximal probability of error

$$R(E, W) = C(E, W) \stackrel{\Delta}{=} \overline{\lim}_{N \rightarrow \infty} \frac{1}{N} \log M(E, N, W),$$

and $\bar{R}(E, W)$ for the case of average probability of error. As in the case of capacity, **E -capacity** is called **maximal** or **average** and denoted, correspondingly, $C(E, W)$ or $\bar{C}(E, W)$ depending on which error probability is considered in (3).

The concept of E -capacity was first considered by the author in [1], where derivation of the upper bound $R_{sp}(E, W)$ was stated. The simple combinatorial proof of $R_{sp}(E, W)$

was obtained in [2]. In Section 4 for completeness we present it, because the paper [2] is little-known.

Alternative methods for the existence part of coding theorems demonstration are Shannon's random coding and Wolfowitz's maximal code methods. In [3] Csiszár and Körner introduced a new original method, based on the lemma of Lovász on graph decomposition. Different methods of error exponent investigation were presented in [4] – [13] and in many other works. Here we shall derive upper bounds for $R(E, W)$ using the method of graph decomposition.

II. FORMULATION OF RESULTS

In the beginning we remind our notations for necessary characteristics of Shannon's entropy and mutual information and Kullback-Leibler's divergence.

The size of the set \mathcal{X} is denoted by $|\mathcal{X}|$. Let P be a PD of RV X

$$P = \{P(x), x \in \mathcal{X}\},$$

V be a conditional PD of RV Y for given value x of RV Y

$$V = \{V(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}.$$

The joint PD of RVs X and Y is

$$P \circ V = \{P \circ V(x, y) = P(x)V(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\},$$

and PD of RV Y is

$$PV = \{PV(y) = \sum_{x \in \mathcal{X}} P(x)V(y|x), y \in \mathcal{Y}\}.$$

Sometimes we need to consider a stochastic matrix $\overline{V} : \mathcal{X} \rightarrow \mathcal{X}$ of conditional probabilities $\overline{V} = \{\overline{V}(x|\tilde{x}), x \in \mathcal{X}, \tilde{x} \in \mathcal{X}\}$.

We use the following notations: for entropy of RV X with PD P :

$$H_P(X) \triangleq - \sum_{x \in \mathcal{X}} P(x) \log P(x),$$

for joint entropy of RVs X and Y :

$$H_{P,V}(X, Y) \triangleq - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(x)V(y|x) \log P(x)V(y|x),$$

for conditional entropy of RV Y relative to RV X :

$$H_{P,V}(Y|X) \triangleq - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(x)V(y|x) \log V(y|x),$$

for mutual information of RVs X and Y :

$$I_{P,V}(X \wedge Y) \triangleq - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(x)V(y|x) \log \frac{V(y|x)}{PV(y)},$$

for informational divergence of PD P and PD Q on \mathcal{X} :

$$D(P\|Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)},$$

and for informational conditional divergence of PD $P \circ V$ and PD $P \circ W$ on $\mathcal{X} \times \mathcal{Y}$:

$$D(V\|W|P) \triangleq \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(x)V(y|x) \log \frac{V(y|x)}{W(y|x)}.$$

The following identities are often useful

$$D(P \circ V\|Q \circ W) = D(P\|Q) + D(V\|W|P),$$

$$H_{P,V}(X, Y) = H_P(X) + H_{P,V}(Y|X) = H_{PV}(Y) + H_{P,V}(X|Y),$$

$$I_{P,V}(X \wedge Y) = H_{PV}(Y) - H_{P,V}(Y|X) = H_P(X) + H_{PV}(Y) - H_{P,V}(X, Y).$$

The proofs in this paper will be based on the method of types [9]. The type P of a sequence (or vector) $\mathbf{x} = (x_1, \dots, x_N) \in \mathcal{X}^N$ is a PD $P = \{P(x) = N(x|\mathbf{x})/N, x \in \mathcal{X}\}$, where $N(x|\mathbf{x})$ is the number of repetitions of symbol x among x_1, \dots, x_N . The joint type of \mathbf{x} and $\mathbf{y} \in \mathcal{Y}^N$ is the PD $P = \{P(x, y) = N(x, y|\mathbf{x}, \mathbf{y})/N, x \in \mathcal{X}, y \in \mathcal{Y}\}$, where $N(x, y|\mathbf{x}, \mathbf{y})$ is the number of occurrences of symbols pair (x, y) in the pair of vectors (\mathbf{x}, \mathbf{y}) . We say that the conditional type of \mathbf{y} for given \mathbf{x} is PD $V = \{V(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ if $N(x, y|\mathbf{x}, \mathbf{y}) = N(x|\mathbf{x})V(y|x)$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$.

The set of all PD on \mathcal{X} is denoted by $\mathcal{P}(\mathcal{X})$ and the subset of $\mathcal{P}(\mathcal{X})$ consisting of the possible types of sequences $\mathbf{x} \in \mathcal{X}^N$ is denoted by $\mathcal{P}_N(\mathcal{X})$. The set of vectors \mathbf{x} of type P is denoted by $\mathcal{T}_P^N(X)$ and $\mathcal{T}_P^N(X) = \emptyset$ for PD $P \in \mathcal{P}(\mathcal{X}) - \mathcal{P}_N(\mathcal{X})$. The set of all sequences $\mathbf{y} \in \mathcal{Y}^N$ of conditional type V for given $\mathbf{x} \in \mathcal{T}_P^N(X)$ is denoted by $\mathcal{T}_{P,V}^N(Y|\mathbf{x})$ and called V -shell of \mathbf{x} . The set of all possible V -shells for \mathbf{x} of type P is denoted $\mathcal{V}_N(\mathcal{Y}, P)$.

In the following lemmas very useful properties of types are formulated, for proofs see [9].

Lemma 1: (Type counting)

$$|\mathcal{P}_N(\mathcal{X})| < (N+1)^{|\mathcal{X}|}, \quad (4)$$

$$|\mathcal{V}_N(\mathcal{Y}, P)| < (N+1)^{|\mathcal{X}||\mathcal{Y}|}. \quad (5)$$

Lemma 2: For any type $P \in \mathcal{P}_N(\mathcal{X})$

$$(N+1)^{-|\mathcal{X}|} \exp\{NH_P(X)\} \leq |\mathcal{T}_P^N(X)| \leq \exp\{NH_P(X)\}, \quad (6)$$

and for any conditional type V and $\mathbf{x} \in \mathcal{T}_P^N(X)$

$$(N+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp\{NH_{P,V}(Y|X)\} \leq |\mathcal{T}_{P,V}^N(Y|\mathbf{x})| \leq \exp\{NH_{P,V}(Y|X)\}. \quad (7)$$

Lemma 3: If $\mathbf{x} \in \mathcal{T}_P^N(X)$, then for every PD Q on \mathcal{X}

$$Q^N(\mathbf{x}) = \prod_{n=1}^N Q(x_n) = \exp\{-N(H_P(X) + D(P\|Q))\}. \quad (8)$$

If $\mathbf{y} \in \mathcal{T}_{P,V}^N(Y|\mathbf{x})$, then for every conditional PD V on \mathcal{Y} for given \mathbf{x}

$$W^N(\mathbf{y}|\mathbf{x}) = \exp\{-N(H_{P,V}(Y|X) + D(V\|W|P))\}. \quad (9)$$

Consider the random coding exponent function $R_r(E, W)$, which is a lower estimate for $C(E, W) = R(E, W)$

$$\begin{aligned} R_r(P, E, W) &\stackrel{\Delta}{=} \min_{V:D(V\|W|P)\leq E} |I_{P,V}(X \wedge Y) + D(V\|W|P) - E|^+, \\ R_r(E, W) &\stackrel{\Delta}{=} \max_P R_r(P, E, W). \end{aligned} \quad (10)$$

The expurgated exponent function $R_x(E, W)$, which is another lower estimate for $R(E, W)$ is defined using probability matrix $\bar{V} : \mathcal{X} \rightarrow \mathcal{X}$

$$R_x(P, E, W) = \min_{\bar{V}} \{I_{P,\bar{V}}(X \wedge \bar{X}) + |\mathbf{E}_{P,\bar{V}} d_B(X, \bar{X}) - E|^+\},$$

here $d_B(x, \tilde{x})$ is the Bhattacharyya distance

$$d_B(x, \tilde{x}) \stackrel{\Delta}{=} -\log \sum_{y \in \mathcal{Y}} \sqrt{W(y|x)W(y|\tilde{x})}, \quad (11)$$

and

$$R_x(E, W) \stackrel{\Delta}{=} \max_P R_x(P, E, W).$$

Sphere packing exponent function serves an upper bound of $R(E, W)$.

$$\mathcal{R}_{sp}(P, E, W) = \min_{V:D(V\|W|P)\leq E} I_{P,V}(X \wedge Y), \quad (12)$$

$$R_{sp}(E, W) = \max_P R_{sp}(P, E, W).$$

It was first considered in [1].

Theorem 1: For DMC W and for any $E > 0$ the following bound holds

$$R(E, W) \geq \max(R_r(E, W), R_x(E, W)).$$

Theorem 2: For DMC W for $E > 0$ the following inequalities hold

$$R(E, W) \leq \bar{R}(E, W) \leq R_{sp}(E, W).$$

Theorem 3: For $0 < E \leq E_{cr}(P, W)$, where

$$E_{cr}(P, W) = \min \left\{ E : \frac{\partial R_{sp}(P, E, W)}{\partial E} \geq -1 \right\},$$

the estimates are equal each other and give E -capacity:

$$R(P, E, W) = R_{sp}(P, E, W) = R_r(P, E, W).$$

For

$$E_{cr}(W) = \min \left\{ E : \frac{\partial R_{sp}(E, W)}{\partial E} \geq -1 \right\} = \max_P E_{cr}(P, W)$$

the equality holds:

$$R(E, W) = R_{sp}(E, W) = R_r(E, W).$$

Remark 1: For $E \rightarrow 0$

$$\lim_{E \rightarrow 0} R_{sp}(P, E, W) = \lim_{E \rightarrow 0} R_r(P, E, W) = I_{P,W}(X \wedge Y),$$

$$R_{sp}(0, W) = R_r(0, W) = C(W).$$

III. PROOF OF THEOREM 1

Lemma 4: Consider a finite set \mathcal{A} and a nonnegative function ν on $\mathcal{A} \times \mathcal{A}$ such that for every $a, b \in \mathcal{A}$

$$\nu(a, b) = \nu(b, a), \quad \nu(a, a) = 0.$$

If for some t , for each $a \in \mathcal{A}$

$$\sum_{b \in \mathcal{A}} \nu(a, b) < t,$$

and t_1, t_2, \dots, t_S are nonnegative numbers such that

$$\sum_{s=1}^S t_s \geq t,$$

then \mathcal{A} can be divided into S disjoint subsets $\mathcal{A}_1, \dots, \mathcal{A}_S$ such that for every $a \in \mathcal{A}_s$

$$\sum_{b \in \mathcal{A}_s} \nu(a, b) < t_s, \quad s = \overline{1, S}.$$

For proof of the Lemma see [3]. There lower bounds for reliability function $E(R, W)$ of DMC and of sources with side information were obtained using Lemma 4. We now present similar derivation of random coding and expurgated bounds for E -capacity $R(E, W)$ of DMC.

Theorem 1 formulated above is a consequence of the following

Theorem 4: For DMC $W : \mathcal{X} \rightarrow \mathcal{Y}$, any $E > 0$, $\delta > 0$ and type $P \in \mathcal{P}_N(\mathcal{X})$ for sufficiently large N codes (f, g) exist such that

$$e(f, g, N, W) = \exp\{-N(E + \delta)\} \tag{13}$$

and

$$R(P, f, g, N) \geq \max(R_r(P, E + \delta, W), R_x(P, E + \delta, W)).$$

The proof of Theorem 4 consists of several steps. First we shall prove

Lemma 5: For given type $P \in \mathcal{P}_N(\mathcal{X})$, for any $0 < r < |\mathcal{T}_P^N(X)|$ such set \mathcal{C} exists, that $\mathcal{C} \subset \mathcal{T}_P^N(X)$, $|\mathcal{C}| \geq r$ and for any $\tilde{\mathbf{x}} \in \mathcal{C}$ and matrix $\overline{V} : \mathcal{X} \rightarrow \mathcal{X}$ different from the identity matrix the following inequality holds

$$|\mathcal{T}_{P, \overline{V}}^N(X | \tilde{\mathbf{x}}) \cap \mathcal{C}| \leq r |\mathcal{T}_{P, \overline{V}}^N(X | \tilde{\mathbf{x}})| \exp\{-N(H_P(X) - \delta_N)\}, \tag{14}$$

where

$$\delta_N = N^{-1}[(|\mathcal{X}|^2 + |\mathcal{X}|) \log(N+1) + 1].$$

Proof: Using Lemma 4 let us assume $\mathcal{A} \triangleq \mathcal{T}_P^N(X)$ and

$$\nu(\mathbf{x}, \tilde{\mathbf{x}}) \triangleq \begin{cases} |\mathcal{T}_{P,\overline{V}}^N(X|\tilde{\mathbf{x}})|^{-1}, & \text{if } \mathbf{x} \neq \tilde{\mathbf{x}} \text{ and } \mathbf{x} \in \mathcal{T}_{P,\overline{V}}^N(X|\tilde{\mathbf{x}}), \\ 0, & \text{if } \mathbf{x} = \tilde{\mathbf{x}}. \end{cases}$$

Because \mathbf{x} and $\tilde{\mathbf{x}}$ are of the same type P , when $\mathbf{x} \in \mathcal{T}_{P,\overline{V}}^N(X|\tilde{\mathbf{x}})$, then $\tilde{\mathbf{x}} \in \mathcal{T}_{P,\overline{V}'}^N(X|\mathbf{x})$ where \overline{V}' is such that the matrix with elements $P(\tilde{x})\overline{V}(x|\tilde{x})$ is transposed to the matrix with elements $P(x)\overline{V}'(\tilde{x}|x)$, $x, \tilde{x} \in \mathcal{X}$. Here $NP(\tilde{x})\overline{V}(x|\tilde{x}) = N(x, \tilde{x}|\mathbf{x}, \tilde{\mathbf{x}})$ and $NP(x)\overline{V}'(\tilde{x}|x) = N(\tilde{x}, x|\tilde{\mathbf{x}}, \mathbf{x})$ define correspondingly joint types of the pairs $(\mathbf{x}, \tilde{\mathbf{x}})$ and $(\tilde{\mathbf{x}}, \mathbf{x})$. We have that

$$|\mathcal{T}_{P,\overline{V}}^N(X|\tilde{\mathbf{x}})| = \prod_{\tilde{x}} \frac{(NP(\tilde{x}))!}{\prod_x (N(\tilde{x}, x|\tilde{\mathbf{x}}, \mathbf{x}))!}$$

and

$$|\mathcal{T}_{P,\overline{V}}^N(X|\mathbf{x})| = \prod_{\tilde{x}} \frac{(NP(\tilde{x}))!}{\prod_x (N(x, \tilde{x}|\mathbf{x}, \tilde{\mathbf{x}}))!},$$

the right sides are equal, so we see that $\nu(\mathbf{x}, \tilde{\mathbf{x}}) = \nu(\tilde{\mathbf{x}}, \mathbf{x})$. We have also from (5)

$$\sum_{\mathbf{x} \in \mathcal{T}_P^N(X)} \nu(\tilde{\mathbf{x}}, \mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{T}_P^N(X)} \nu(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{\overline{V}} \sum_{\mathbf{x} \in \mathcal{T}_{P,\overline{V}}^N(X|\tilde{\mathbf{x}})} \nu(\mathbf{x}, \tilde{\mathbf{x}}) < (N+1)^{|\mathcal{X}|^2}.$$

If we take $t \triangleq (N+1)^{|\mathcal{X}|^2}$, $t_s \triangleq t/S$, $s = \overline{1, S}$, then according to Lemma 4 there exists a partition of $\mathcal{T}_P^N(X)$ into subsets \mathcal{A}_s , $s = \overline{1, S}$, such that for each $\tilde{\mathbf{x}}$ from \mathcal{A}_s

$$|\mathcal{T}_{P,\overline{V}}^N(X|\tilde{\mathbf{x}}) \cap \mathcal{A}_s| \leq \frac{1}{S} |\mathcal{T}_{P,\overline{V}}^N(X|\tilde{\mathbf{x}})| (N+1)^{|\mathcal{X}|^2}, \quad s = \overline{1, S}. \quad (15)$$

Taking \mathcal{C} equal to greatest \mathcal{A}_s and S equal to integer part of $|\mathcal{T}_P^N(X)|/r$ we receive $|\mathcal{C}| \geq S^{-1}|\mathcal{T}_P^N(X)| \geq r$, and inequality (14), which follows from (15) and (6), because

$$\frac{1}{S} = \frac{1}{\lfloor |\mathcal{T}_P^N(X)|/r \rfloor} \leq \frac{1}{|\mathcal{T}_P^N(X)|/2r} = \frac{2r}{|\mathcal{T}_P^N(X)|}.$$

Lemma 5 is proved.

For code existence theorems demonstration it is possible to consider various "good" decoding rules. For definition of those rules following [3], we apply different real-valued functions α defined on $\mathcal{X}^N \times \mathcal{Y}^N$. One says that g_α decoding is used if to each \mathbf{y} from \mathcal{Y}^N on the output of the channel the message m is accepted when codeword $\mathbf{x}(m)$ minimizes $\alpha(\mathbf{x}(m), \mathbf{y})$. One uses such functions α which depend only on type P of \mathbf{x} and conditional type V of \mathbf{y} for given \mathbf{x} . Such functions α can be written in the form $\alpha(P, V)$ and at respective decoding

$$g_\alpha : \mathcal{Y}^N \rightarrow \mathcal{M},$$

the message m' corresponds to the vector \mathbf{y} , if

$$(m', \tilde{V}) = \arg \min_{(m, V) : \mathbf{y} \in \mathcal{T}_{P,V}^N(Y|f(m))} \alpha(P, V).$$

Here $\tilde{V} = \{\tilde{V}(y|\tilde{x}), \tilde{x} \in \mathcal{X}, y \in \mathcal{Y}\}$ is a matrix different from V but guaranteeing that

$$\sum_{x \in \mathcal{X}} P(x)V(y|x) = \sum_{\tilde{x} \in \mathcal{X}} P(\tilde{x})\tilde{V}(y|\tilde{x}), \quad y \in \mathcal{Y},$$

or equivalently, $PV = P\tilde{V}$.

Previously the following two decoding rules were used [3]: **maximum-likelihood decoding**, when the accepted codeword $\mathbf{x}(m)$ maximizes the transition probability $W^N(\mathbf{y}|\mathbf{x}(m))$, in this case according to (9)

$$\alpha(P, V) = D(V\|W|P) + H_{P,V}(Y|X), \quad (16)$$

and the second decoding rule, called **minimum-entropy decoding**, according to which the codeword $\mathbf{x}(m)$ minimizing $H_{P,V}(Y|X)$ is accepted, that is

$$\alpha(P, V) = H_{P,V}(Y|X). \quad (17)$$

In [11] and [12] it was proposed another decoding rule by minimization of

$$\alpha(P, V) = D(V\|W|P), \quad (18)$$

which can be called **minimum - divergence decoding**.

Let $\tilde{\tilde{V}} = \{\tilde{\tilde{V}}(y|x, \tilde{x}), x \in \mathcal{X}, \tilde{x} \in \mathcal{X}, y \in \mathcal{Y}\}$, be a conditional distribution of Y given values of X and \tilde{X} such, that

$$\sum_{\tilde{x}} P(\tilde{x})\bar{V}(x|\tilde{x})\tilde{\tilde{V}}(y|x, \tilde{x}) = P(x)V(y|x), \quad x \in \mathcal{X}, y \in \mathcal{Y}, \quad (19)$$

$$\sum_x P(x)\bar{V}'(\tilde{x}|x)\tilde{\tilde{V}}(y|x, \tilde{x}) = P(\tilde{x})\tilde{V}(y|\tilde{x}), \quad \tilde{x} \in \mathcal{X}, y \in \mathcal{Y}. \quad (20)$$

Using the notation from [3] we write

$$\tilde{V} \prec_\alpha V \text{ if } \alpha(P, \tilde{V}) \leq \alpha(P, V) \text{ and } P\tilde{V} = PV.$$

Let us denote

$$R_\alpha(P, E, W) \triangleq \min_{\substack{\tilde{V}, \tilde{\tilde{V}}, \tilde{V}: \tilde{V} \prec_\alpha V, V: D(V\|W|P) \leq E}} \{I_{P, \bar{V}}(X \wedge \tilde{X}) + |I_{P, \bar{V}, \tilde{\tilde{V}}}(Y \wedge \tilde{X}|X) + D(V\|W|P) - E|^+\}, \quad (21)$$

where RV X, \tilde{X}, Y have values, correspondingly, on $\mathcal{X}, \mathcal{X}, \mathcal{Y}$ such that the following is valid:
both X and \tilde{X} have distribution P and $P\bar{V} = P$,

$\tilde{\tilde{V}}$ is the conditional distribution of RV Y given X and \tilde{X} satisfying (19) and (20).

Minimization in (21) (and later) is understud by variables ordered from right to left. In (21) min must be taken by V under condition $D(V\|W|P) \leq E$, then by \tilde{V} under condition $\tilde{V} \prec_\alpha V$, by $\tilde{\tilde{V}}$ under conditions (19), (20) and at last by \bar{V} .

The main portion of Theorem 4 demonstration is contained in

Proposition 1: For any DMC W , any type $P \in \mathcal{P}_N(\mathcal{X})$, any $E > 0$, $\delta_N > 0$, for all sufficiently large N codes (f, g_α) exist such, that

$$e(f, g_\alpha, N, W) = \exp\{-N(E + \delta_N/2)\}, \quad (22)$$

and the rate is large enough:

$$R(P, f, g_\alpha, N) \geq R_\alpha(P, E + \delta_N, W). \quad (23)$$

Proof: For some R let us write $r = \exp\{N(R - \delta_N)\}$. According to Lemma 5 for $r < |\mathcal{T}_P^N(X)|$ a collection $\mathcal{C} \subset \mathcal{T}_P^N(X)$ exists such that $|\mathcal{C}| \geq r$ and for each $\tilde{\mathbf{x}} \in \mathcal{C}$ and any probability matrix $\bar{V} : \mathcal{X} \rightarrow \mathcal{X}$ different from the identity matrix, for N large enough

$$|\mathcal{T}_{P,\bar{V}}^N(X|\tilde{\mathbf{x}}) \cap \mathcal{C}| \leq \exp\{N(R - I_{P,\bar{V}}(X \wedge \tilde{X}))\}. \quad (24)$$

Remark that from symmetry we have

$$\begin{aligned} |\mathcal{T}_{P,\bar{V}'}^N(\tilde{X}|\mathbf{x}) \cap \mathcal{C}| &\leq \exp\{N(R - I_{P,\bar{V}'}(\tilde{X} \wedge X))\} = \\ &= \exp\{N(R - I_{P,\bar{V}}(X \wedge \tilde{X}))\}. \end{aligned} \quad (25)$$

Let us take \mathcal{C} as a set of codewords of the code (f, g_α) . If $\mathbf{x} \in \mathcal{T}_{P,\bar{V}}^N(X|\tilde{\mathbf{x}})$ exists such that $\mathbf{x} \in \mathcal{C}$ and $\tilde{\mathbf{x}} \in \mathcal{C}$, then

$$R(P, f, g_\alpha, N) \geq I_{P,\bar{V}}(X \wedge \tilde{X}). \quad (26)$$

As in (1)

$$e(m, f, g_\alpha, N, W) = W^N(\mathcal{Y}^N - g_\alpha^{-1}(m)|f(m)).$$

In accordance with g_α -decoding the set $\mathcal{Y}^N - g_\alpha^{-1}(m)$ contains all words \mathbf{y} for which codeword $f(m')$ exists different from $f(m)$, such that, if $f(m) \in \mathcal{T}_{P,\bar{V}}^N(X|f(m'))$ and $\mathbf{y} \in \mathcal{T}_{P,\bar{V}}^N(Y|f(m)) \cap \mathcal{T}_{P,\tilde{V}}^N(Y|f(m'))$, then the joint type of the triple $(f(m), f(m'), \mathbf{y})$ is $P \circ \bar{V} \circ \tilde{V}$, with \tilde{V} meeting conditions (19), (20) and $\tilde{V} \prec_\alpha V$.

Denote $(\mathcal{Y}^N - g_\alpha^{-1}(m)) \cap \mathcal{T}_{P,\bar{V},\tilde{V}}^N(Y|f(m), f(m'))$ the set of such vectors \mathbf{y} , for which the triple $(f(m), f(m'), \mathbf{y})$ has that joint type $P \circ \bar{V} \circ \tilde{V}$ for some m' different from m . According to (5) the number of such types does not exceed $(N+1)^{|\mathcal{X}|^2|\mathcal{Y}|}$. Taking into account that

$$\mathcal{T}_{P,\bar{V},\tilde{V}}^N(Y|f(m), f(m')) \subset \mathcal{T}_{P,\bar{V}}^N(Y|f(m)), \quad (27)$$

and then using (9) we have

$$\begin{aligned} W^N(\mathcal{Y}^N - g_\alpha^{-1}(m)|f(m)) &\leq (N+1)^{|\mathcal{X}|^2|\mathcal{Y}|} \times \\ &\times \max_{\bar{V}, \tilde{V}, \tilde{V}: \tilde{V} \prec_\alpha V, V} \sum_{f(m'): f(m) \in \mathcal{T}_{P,\bar{V}}^N(X|f(m')) \cap \mathcal{C}} |(\mathcal{Y}^N - g_\alpha^{-1}(m)) \cap \mathcal{T}_{P,\bar{V},\tilde{V}}^N(Y|f(m), f(m'))| W^N(\mathbf{y}|f(m)) = \\ &= (N+1)^{|\mathcal{X}|^2|\mathcal{Y}|} \max_{\bar{V}, \tilde{V}, \tilde{V}: \tilde{V} \prec_\alpha V, V} \sum_{f(m'): f(m) \in \mathcal{T}_{P,\bar{V}}^N(X|f(m')) \cap \mathcal{C}} |(\mathcal{Y}^N - g_\alpha^{-1}(m)) \cap \mathcal{T}_{P,\bar{V},\tilde{V}}^N(Y|f(m), f(m'))| \times \\ &\quad \times \exp\{-N(D(V||W|P) + H_{P,V}(Y|X))\}. \end{aligned} \quad (28)$$

Granting (2), (7) and (24) we bound

$$\sum_{f(m'): f(m) \in \mathcal{T}_{P,\bar{V}}^N(X|f(m')) \cap \mathcal{C}} |(\mathcal{Y}^N - g_\alpha^{-1}(m)) \cap \mathcal{T}_{P,\bar{V},\tilde{V}}^N(Y|f(m), f(m'))| \leq$$

$$\begin{aligned}
&\leq \exp\{N(R(P, f, g_\alpha, N) - I_{P, \bar{V}}(X \wedge \bar{X}))\} \exp\{NH_{P, \bar{V}, \tilde{V}}(Y|X, \bar{X})\} = \\
&= \exp\{N(H_{P, V}(Y|X) - I_{P, \bar{V}, \tilde{V}}(X, Y \wedge \bar{X}) + R(P, f, g_\alpha, N))\}.
\end{aligned} \tag{29}$$

Since (27) is valid, from (7) we have

$$\sum_{f(m'): f(m) \in \mathcal{T}_{P, \bar{V}}^N(X|f(m')) \cap C} |(\mathcal{Y}^N - g_\alpha^{-1}(m)) \cap \mathcal{T}_{P, \bar{V}, \tilde{V}}^N(Y|f(m), f(m'))| \leq \exp\{NH_{P, V}(Y|X)\}.$$

With (29) it gives us

$$\begin{aligned}
&\sum_{f(m'): f(m) \in \mathcal{T}_{P, \bar{V}}^N(X|f(m')) \cap C} |(\mathcal{Y}^N - g_\alpha^{-1}(m)) \cap \mathcal{T}_{P, \bar{V}, \tilde{V}}^N(Y|f(m), f(m'))| \leq \\
&\leq \exp\{N(H_{P, V}(Y|X) - |I_{P, \bar{V}, \tilde{V}}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N)|^+)\}.
\end{aligned} \tag{30}$$

From (28) and (30) we obtain

$$\begin{aligned}
e(f, g_\alpha, N, W) &= \max_m W^N(\mathcal{Y}^N - g_\alpha^{-1}(m)|f(m)) \leq \\
&\leq (N+1)^{|\mathcal{X}|^2|\mathcal{Y}|} \max_{\bar{V}, \tilde{V}, \tilde{V}: \tilde{V} \prec_\alpha V, V} \exp\{-N(D(V\|W|P) + |I_{P, \bar{V}, \tilde{V}}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N)|^+)\}.
\end{aligned}$$

If the equality in (22) is in force, then for sufficiently large N , we have

$$0 \geq \min_{\bar{V}, \tilde{V}, \tilde{V}: \tilde{V} \prec_\alpha V, V} (D(V\|W|P) + |I_{P, \bar{V}, \tilde{V}}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N)|^+ - E - \delta_N). \tag{31}$$

Suppose that this minimum is obtained on \tilde{V}_0 with V_0 and \bar{V}_0 , that is

$$0 \geq D(V_0\|W|P) - E - \delta_N + |I_{P, \bar{V}_0, \tilde{V}_0}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N)|^+.$$

Because $|I_{P, \bar{V}_0, \tilde{V}_0}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N)|^+ \geq 0$, we have

$$\begin{aligned}
0 &\geq \min_{\bar{V}, \tilde{V}, \tilde{V}: \tilde{V} \prec_\alpha V, V: D(V\|W|P) \leq E + \delta_N - |I_{P, \bar{V}_0, \tilde{V}_0}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N)|^+} (D(V\|W|P) - E - \delta_N + \\
&\quad + |I_{P, \bar{V}, \tilde{V}}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N)|^+) \geq \\
&\geq \min_{\bar{V}, \tilde{V}, \tilde{V}: \tilde{V} \prec_\alpha V, V: D(V\|W|P) \leq E + \delta_N} (D(V\|W|P) - E - \delta_N + |I_{P, \bar{V}, \tilde{V}}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N)|^+).
\end{aligned}$$

Since

$$|I_{P, \bar{V}, \tilde{V}}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N)|^+ \geq I_{P, \bar{V}, \tilde{V}}(X, Y \wedge \bar{X}) - R(P, f, g_\alpha, N),$$

then we deduce that

$$R(P, f, g_\alpha, N) \geq \min_{\bar{V}, \tilde{V}, \tilde{V}: \tilde{V} \prec_\alpha V, V: D(V\|W|P) \leq E + \delta_N} (D(V\|W|P) - E - \delta_N + |I_{P, \bar{V}, \tilde{V}}(X, Y \wedge \bar{X})|). \tag{32}$$

From (32), (21) and (26) we obtain (23). Proposition 1 is proved.

Remark 2: The functions $R_\alpha(P, E, W)$ and $R_x(P, E, W)$ depend on E continuously.

Lemma 6: Let us introduce the following functions

$$R_{\alpha,r}(P, E, W) \triangleq \min_{\tilde{V}: \tilde{V} \prec_\alpha V, V: D(V||W|P) \leq E} |I_{P,\tilde{V}}(Y \wedge \tilde{X}) + D(V||W|P) - E|^+, \quad (33)$$

$$R_{\alpha,x}(P, E, W) \triangleq \min_{\overline{V}, \tilde{\overline{V}}, \tilde{V}: \tilde{V} \prec_\alpha V, V} \{I_{P,\overline{V}}(X \wedge \tilde{X}) + |I_{P,\overline{V},\tilde{V}}(Y \wedge \tilde{X}|X) + D(V||W|P) - E|^+\}. \quad (34)$$

Then for all P and $E > 0$

$$R_\alpha(P, E, W) \geq \max[R_{\alpha,x}(P, E, W), R_{\alpha,r}(P, E, W)].$$

Proof: The inequality

$$R_\alpha(P, E, W) \geq R_{\alpha,x}(P, E, W)$$

follows from definitions (21) and (34). For the proof of the inequality

$$R_\alpha(P, E, W) \geq R_{\alpha,r}(P, E, W)$$

remark that

$$I_{P,\overline{V}}(X \wedge \tilde{X}) + I_{P,\overline{V},\tilde{V}}(Y \wedge \tilde{X}|X) = I_{P,\overline{V},\tilde{V}}(XY \wedge \tilde{X}) \geq I_{P,\tilde{V}}(Y \wedge \tilde{X})$$

and then compare (21) and (33) using inequality $|a + b|^+ \leq |a|^+ + |b|^+$, which is valid for any real a and b .

Lemma 7: A point $E_\alpha^*(P, W)$ exists, such that

$$\max[R_{\alpha,x}(P, E, W), R_{\alpha,r}(P, E, W)] = \begin{cases} R_{\alpha,r}(P, E, W), & \text{when } E \leq E_\alpha^*(P, W) \\ R_{\alpha,x}(P, E, W), & \text{when } E \geq E_\alpha^*(P, W). \end{cases}$$

Proof: Note that functions $R_{\alpha,r}(P, E, W)$ and $R_{\alpha,x}(P, E, W)$ are nonnegative and decreasing by E . Let us first prove that for any $E \geq E' \geq 0$

$$R_{\alpha,x}(P, E', W) \leq R_{\alpha,x}(P, E, W) + E - E'. \quad (35)$$

In accordance with (34), bearing in mind the inequality $|a + b|^+ \leq |a|^+ + |b|^+$, we obtain

$$\begin{aligned} R_{\alpha,x}(P, E', W) &= \min_{\overline{V}, \tilde{\overline{V}}, \tilde{V}: \tilde{V} \prec_\alpha V, V} \{I_{P,\overline{V}}(X \wedge \tilde{X}) + |I_{P,\overline{V},\tilde{V}}(Y \wedge \tilde{X}|X) + D(V||W|P) - E' + E - E'|^+\} \leq \\ &\leq R_{\alpha,x}(P, E, W) + E - E'. \end{aligned}$$

Denote by $E_{\alpha,r}^0(P, W)$ the least value of E , for which $R_{\alpha,r}(E, P, W) = 0$. Let us show that for any E and E' , such that

$$0 \leq E' \leq E \leq E_{\alpha,r}^0(P, W),$$

the inequality

$$R_{\alpha,r}(P, E, W) + E - E' \leq R_{\alpha,r}(P, E', W) \quad (36)$$

holds. Really, in the interval $[0, E_{\alpha,r}^0(P, W))$ function $R_{\alpha,r}(E, P, W)$ is strictly positive, then for such E and E'

$$\begin{aligned} R_{\alpha,r}(P, E', W) &= \min_{\tilde{V}: \tilde{V} \prec_{\alpha} V, V: D(V||W|P) \leq E'} (I_{P,\tilde{V}}(Y \wedge \tilde{X}) + D(V||W|P) - E) + E - E' \geq \\ &\geq \min_{\tilde{V}: \tilde{V} \prec_{\alpha} V, V: D(V||W|P) \leq E} (I_{P,\tilde{V}}(Y \wedge \tilde{X}) + D(V||W|P) - E) + E - E' = \\ &= R_{\alpha,r}(P, E, W) + E - E'. \end{aligned}$$

Denote by $E_{\alpha}^*(P, W)$ the smallest E , for which

$$R_{\alpha,r}(P, E, W) \leq R_{\alpha,x}(P, E, W).$$

Let us show that this inequality holds for all E greater than $E_{\alpha}^*(P, W)$. Consider two cases.

If $0 \leq E_{\alpha}^*(P, W) < E_{\alpha,r}^0(P, W)$, then it follows from (33), (34), (35) and (36) that for all E from interval $(E_{\alpha}^*(P, W), E_{\alpha,r}^0(P, W))$

$$\begin{aligned} R_{\alpha,r}(P, E, W) + E - E_{\alpha}^*(P, W) &\leq R_{\alpha,r}(P, E_{\alpha}^*(P, W), W) \leq \\ &\leq R_{\alpha,x}(P, E_{\alpha}^*(P, W), W) \leq R_{\alpha,x}(P, E, W) + E - E_{\alpha}^*(P, W). \end{aligned}$$

If $E_{\alpha,r}^0(P, W) \leq E_{\alpha}^*(P, W)$, then for all E greater than $E_{\alpha}^*(P, W)$ we have

$$R_{\alpha,x}(P, E, W) = 0 = R_{\alpha,r}(P, E, W).$$

In this case $E_{\alpha}^*(P, W) = E_{\alpha,x}^0(P, W)$.

Lemma 8: For each α -decoding defined in (16), (17), or (18)

$$R_{\alpha,x}(P, E, W) \leq R_x(P, E, W), \quad (37)$$

moreover, for maximum likelihood decoding given by (16) the equality holds.

Proof: First we prove the inequality (37). As with (19) we have

$$D(V||W|P) + I_{P,\bar{V},\tilde{V}}(Y \wedge \tilde{X}|X) = \sum_{x,\tilde{x},y} P(x)\bar{V}(\tilde{x}|x)\tilde{V}(y|x, \tilde{x}) \log \frac{\tilde{V}(y|x, \tilde{x})}{W(y|x)}, \quad (38)$$

and by (20)

$$D(\tilde{V}||W|P) + I_{P,\bar{V},\tilde{V}}(Y \wedge X|\tilde{X}) = \sum_{x,\tilde{x},y} P(x)\bar{V}(\tilde{x}|x)\tilde{V}(y|x, \tilde{x}) \log \frac{\tilde{V}(y|x, \tilde{x})}{W(y|\tilde{x})}, \quad (39)$$

the left parts of which are equal when $\tilde{V} = V$, then from (34)

$$\begin{aligned} R_{\alpha,x}(P, E, W) &\leq \min_{\bar{V}, \tilde{V}: \tilde{V} = V} \{I_{\hat{P},\bar{V}}(X \wedge \tilde{X}) + \\ &+ |(1/2)(I_{P,\bar{V},\tilde{V}}(Y \wedge \tilde{X}|X) + D(V||W|P)) + (1/2)(I_{P,\bar{V},\tilde{V}}(Y \wedge \tilde{X}|X) + D(\tilde{V}||W|P)) - E|^+\}. \end{aligned}$$

From (11), (38) and (39) denoting

$$\tilde{V}_1(y|x, \tilde{x}) \triangleq (\exp d_B(x, \tilde{x})) \sqrt{W(y|x)W(y|\tilde{x})} = \frac{\sqrt{W(y|x)W(y|\tilde{x})}}{\sum_{y'} \sqrt{W(y'|x)W(y'|\tilde{x})}}$$

we have

$$\begin{aligned} R_{\alpha,x}(P, E, W) &\leq \min_{\substack{\tilde{V} \\ \overline{V}, \tilde{V}}} (I_{P,\overline{V}}(X \wedge \tilde{X}) + |D(\tilde{V}\|\tilde{V}_1|P) + \mathbf{E}_{P,\overline{V}}d_B(X, \tilde{X}) - E|^+) = \\ &= \min_{\overline{V}} (I_{P,\overline{V}}(X \wedge \tilde{X}) + |\mathbf{E}_{P,\overline{V}}d_B(X, \tilde{X}) - E|^+) = R_x(P, E, W). \end{aligned}$$

Let us now prove that in the case of maximum likelihood decoding

$$R_x(P, E, W) = R_{\alpha,x}(P, E, W). \quad (40)$$

From the condition $\tilde{V} \prec_\alpha V$ and from (16) we have

$$D(\tilde{V}\|W|P) + H_{P,\tilde{V}}(Y|\tilde{X}) \leq D(V\|W|P) + H_{P,V}(Y|X).$$

In accordance with (38), (39) and the last inequality we can deduce

$$D(V\|W|P) + I_{P,\overline{V},\tilde{V}}(Y \wedge \hat{X}|X) \geq D(\tilde{V}\|\tilde{V}_1|P \circ \overline{V}') + \mathbf{E}_{P,\overline{V}}d_B(X, \tilde{X}) \geq \mathbf{E}_{P,\overline{V}}d_B(X, \tilde{X}).$$

Hence

$$\begin{aligned} |I_{P,\overline{V}}(X \wedge \tilde{X}) + I_{P,\overline{V},\tilde{V}}(Y \wedge \tilde{X}|X) + D(V\|W|P) - E|^+ &\geq \\ &\geq |I_{P,\overline{V}}(X \wedge \tilde{X}) + \mathbf{E}_{P,\overline{V}}d_B(X, \tilde{X}) - E|^+, \end{aligned}$$

which is equivalent to inverse inequality to (37) and therefore (40) holds.

Lemma 9: For each α -decoding

$$R_{\alpha,r}(P, E, W) \leq R_r(P, E, W), \quad (41)$$

moreover, for

- maximum likelihood decoding,
- minimum entropy decoding,
- minimum divergence decoding

the equality holds.

Proof: The inequality (41) is valid because (see (33) and (10))

$$R_{\alpha,r}(P, E, W) \leq \min_{\tilde{V}: \tilde{V}=V, V: D(V\|W|P) \leq E} |I_{P,\tilde{V}}(Y \wedge \tilde{X}) + D(V\|W|P) - E|^+ = R_r(P, E, W).$$

For the case of maximal likelihood decoding (16)

$$\tilde{V} \prec_\alpha V \iff \begin{cases} D(\tilde{V}\|W|P) + H_{P,\tilde{V}}(Y|\tilde{X}) \leq D(V\|W|P) + H_{P,V}(Y|X), \\ P\tilde{V} = PV, \end{cases}$$

hence

$$D(\tilde{V}\|W|P) + I_{P,V}(X \wedge Y) \leq D(V\|W|P) + I_{P,\tilde{V}}(\tilde{X} \wedge Y). \quad (42)$$

Denote for brevity

$$\mathcal{G} \triangleq \{(V, \tilde{V}) : P\tilde{V} = PV, D(V\|W|P) \leq E \text{ and (42) holds}\}.$$

Thus, we can write instead of (33)

$$R_{\alpha,r}(P, E, W) = \min_{(V, \tilde{V}) \in \mathcal{G}} |I_{P,\tilde{V}}(\tilde{X} \wedge Y) + D(V\|W|P) - E|^+ \triangleq \min\{B_1, B_2\}, \quad (43)$$

where

$$\begin{aligned} B_1 &\triangleq \min_{(V, \tilde{V}) \in \mathcal{G}, \tilde{V}: D(\tilde{V}\|W|P) \leq D(V\|W|P)} |I_{P,\tilde{V}}(\tilde{X} \wedge Y) + D(V\|W|P) - E|^+, \\ B_2 &\triangleq \min_{(V, \tilde{V}) \in \mathcal{G}, \tilde{V}: D(\tilde{V}\|W|P) \geq D(V\|W|P)} |I_{P,\tilde{V}}(\tilde{X} \wedge Y) + D(V\|W|P) - E|^+. \end{aligned}$$

In turn neither B_1 nor B_2 are not less than $R_r(P, E, W)$. Really, if inequalities (42) and $D(\tilde{V}\|W|P) \geq D(V\|W|P)$ hold simultaneously, then $I_{P,V}(X \wedge Y) \leq I_{P,\tilde{V}}(\tilde{X} \wedge Y)$. Hence

$$B_2 \geq \min_{(V, \tilde{V}) \in \mathcal{G}, \tilde{V}: D(\tilde{V}\|W|P) \geq D(V\|W|P)} |I_{P,V}(X \wedge Y) + D(V\|W|P) - E|^+ = R_r(P, E, W).$$

But

$$\begin{aligned} B_1 &\geq \min_{(V, \tilde{V}) \in \mathcal{G}, \tilde{V}: D(\tilde{V}\|W|P) \leq D(V\|W|P)} |I_{P,\tilde{V}}(\tilde{X} \wedge Y) + D(\tilde{V}\|W|P) - E|^+ = \\ &= \min_{\tilde{V}: D(\tilde{V}\|W|P) \leq E} |I_{P,\tilde{V}}(\tilde{X} \wedge Y) + D(\tilde{V}\|W|P) - E|^+ = R_r(P, E, W). \end{aligned}$$

The two latest inequalities along with (41) and (43) provide the statement of the Lemma for the method of maximum-likelihood decoding.

For the case of minimum-divergence decoding (18)

$$\tilde{V} \prec_{\alpha} V \iff \{D(\tilde{V}\|W|P) \leq D(V\|W|P), P\tilde{V} = PV\}.$$

Then according to (33) and (10)

$$R_{\alpha,r}(P, E, W) \geq \min_{\tilde{V}: D(\tilde{V}\|W|P) \leq E} |I_{P,\tilde{V}}(X \wedge \tilde{Y}) + D(\tilde{V}\|W|P) - E|^+ = R_r(P, E, W).$$

For the case of minimum-entropy decoding from (17)

$$\tilde{V} \prec_{\alpha} V \iff \{H_{P,\tilde{V}}(Y|\tilde{X}) \leq H_{P,V}(Y|X), PV = P\tilde{V}\},$$

which implies $I_{P,\tilde{V}}(\tilde{X} \wedge Y) \geq I_{P,V}(X \wedge Y)$ from where

$$R_{\alpha,r}(P, E, W) \geq \min_{V: D(V\|W|P) \leq E} |I_{P,V}(X \wedge Y) + D(V\|W|P) - E|^+ = R_r(P, E, W).$$

Thus the proof of Theorem 1 is completed by unification of results of Lemmas 4 – 9 and Proposition 1.

IV. PROOF OF THEOREM 2

Let E and δ be given such that $E > \delta > 0$. Let the code (f, g) of length N be defined, R be the rate of the code and average error probability satisfies the analog of condition (3) for $E - \delta$

$$\bar{e}(f, g, N, W) = \exp\{-N(E - \delta)\},$$

which according to Definition 3 is

$$\frac{1}{M} \sum_m W^N \{ \mathcal{T}_P^N - g^{-1}(m) | f(m) \} = \exp\{-N(E - \delta)\}. \quad (44)$$

As the number of messages M may be presented by sum of numbers of codewords of different types $M = \sum_P |f(\mathcal{M}) \cap \mathcal{T}_P^N(X)|$, and the number of all types $P \in \mathcal{P}_N(\mathcal{X})$ is less than $(N + 1)^{|\mathcal{X}|}$ (see (4)), then there exists a "major" type P^* such, that

$$|f(\mathcal{M}) \cap \mathcal{T}_{P^*}^N(X)| \geq M(N + 1)^{-|\mathcal{X}|}. \quad (45)$$

Now for any conditional type V in the left part of (44) we can consider only codewords of type P^* and the part of output vectors \mathbf{y} of the conditional type V

$$\sum_{m: f(m) \in \mathcal{T}_{P^*}^N(X)} W^N \{ \mathcal{T}_{P^*, V}^N(Y | f(m)) - g^{-1}(m) | f(m) \} \leq M \exp\{-N(E - \delta)\}.$$

or

$$\begin{aligned} \sum_{m: f(m) \in \mathcal{T}_{P^*}^N(X)} \{ |\mathcal{T}_{P^*, V}^N(Y | f(m))| - |\mathcal{T}_{P^*, V}^N(Y | f(m)) \cap g^{-1}(m)| \} W^N(\mathbf{y} | \mathbf{x}) &\leq \\ &\leq M \exp\{-N(E - \delta)\}. \end{aligned}$$

From (9) we obtain

$$\begin{aligned} \sum_{m: f(m) \in \mathcal{T}_{P^*}^N(X)} |\mathcal{T}_{P^*, V}^N(Y | f(m))| - \frac{M \exp\{-N(E - \delta)\}}{\exp\{-N(D(V \| W | P^*) + H_{P^*, V}(Y | X))\}} &\leq \\ \leq \sum_{m: f(m) \in \mathcal{T}_{P^*}^N(X)} |\mathcal{T}_{P^*, V}^N(Y | f(m)) \cap g^{-1}(m)|. \end{aligned}$$

It follows from the definition of decoding function g that the sets $g^{-1}(m)$ are disjoint, therefore

$$\sum_{m: f(m) \in \mathcal{T}_{P^*}^N(X)} |\mathcal{T}_{P^*, V}^N(Y | f(m)) \cap g^{-1}(m)| \leq |\mathcal{T}_{P^*, V}^N(Y)|.$$

Then from (7) we have

$$\begin{aligned} |f(\mathcal{M}) \cap \mathcal{T}_{P^*}^N(X)| (N + 1)^{-|\mathcal{X}||\mathcal{Y}|} \exp\{NH_{P^*, V}(Y | X)\} - \\ - M \exp\{N(D(V \| W | P^*) + H_{P^*, V}(Y | X) - E + \delta)\} \leq \exp\{NH_{P^*, V}(Y)\}. \end{aligned}$$

Taking into account (45) we come to an estimate:

$$M \leq \frac{\exp\{NI_{P^*, V}(X \wedge Y)\}}{(N + 1)^{-|\mathcal{X}||\mathcal{Y}|+1} - \exp(N(D(V \| W | P^*) - E + \delta))}.$$

The right part of this inequality can be minimized by the choice of conditional type V , keeping the denominator positive, which takes place for large N when the following inequality holds:

$$D(V\|W|P^*) \leq E - \delta.$$

The statement of Theorem 2 follows from the definitions of $\overline{R}(E, W)$ and $R_{sp}(E, W)$ and from the continuity by E of the function $R_{sp}(P, E, W)$.

Similarly the same bound in the case of maximal error probability can be proved, but it follows also from the given proof.

Example. We shall calculate $R_{sp}(E, W)$ for the binary symmetric channel (BSC). Consider BSC W with

$$\mathcal{X} = \{0, 1\}, \quad \mathcal{Y} = \{0', 1'\},$$

$$W(0'|1) = W(1'|0) = w_1 > 0, \quad W(0'|0) = W(1'|1) = w_2 > 0.$$

Correspondingly, for another BSC V on the same \mathcal{X} and \mathcal{Y} we denote

$$V(0'|1) = V(1'|0) = v_1, \quad V(0'|0) = V(1'|1) = v_2.$$

It is clear that $w_1 + w_2 = 1$, $v_1 + v_2 = 1$.

The maximal value of the mutual information $I_{P,V}(X \wedge Y)$ in the definition of $R_{sp}(E, W)$ comes out on $p^*(0) = p^*(1) = 1/2$ because of symmetry of the channel, therefore

$$I_{P^*,V}(X \wedge Y) = 1 + v_1 \log v_1 + v_2 \log v_2.$$

The condition $D(V\|W|P^*) \leq E$ will take the following form

$$v_1 \log \frac{v_1}{w_1} + v_2 \log \frac{v_2}{w_2} \leq E.$$

So, the problem of extremum with restrictions must be solved (see (12)):

$$\begin{cases} -(1 + v_1 \log v_1 + v_2 \log v_2) = \max \\ -(v_1 \log \frac{v_1}{w_1} + v_2 \log \frac{v_2}{w_2} - E) = 0 \\ v_1 + v_2 = 1. \end{cases}$$

Using Kuhn-Takker theorem, we find that \bar{v}_1, \bar{v}_2 give the solution of the problem if and only if there exist $\lambda_1 > 0, \lambda_2 > 0$, satisfying the following conditions

$$\begin{cases} \frac{\partial}{\partial v_i} (-1 - v_1 \log v_1 - v_2 \log v_2) + \lambda_1 \frac{\partial}{\partial v_i} (-v_1 \log \frac{v_1}{w_1} - v_2 \log \frac{v_2}{w_2} + E) + \\ + \lambda_2 \frac{\partial}{\partial v_i} (v_1 + v_2 - 1) = 0, \quad i = 1, 2, \\ \lambda_1 (v_1 \log \frac{v_1}{w_1} + v_2 \log \frac{v_2}{w_2} - E) = 0, \end{cases}$$

which for \bar{v}_1 and \bar{v}_2 giving maximum are equivalent to

$$\begin{cases} \log \bar{v}_i + \log e = -\lambda_1 (\log \frac{\bar{v}_i}{w_i} + \log e + \lambda_2), \quad i = 1, 2, \\ \bar{v}_1 \log \frac{\bar{v}_1}{w_1} + \bar{v}_2 \log \frac{\bar{v}_2}{w_2} = E. \end{cases} \quad (46)$$

Solving the first two equations from (46) we obtain

$$\bar{v}_i = w_i^{\frac{\lambda_1}{1+\lambda_1}} 2^{-\frac{1}{1+\lambda_1}(\lambda_1 - \lambda_2 + 1) \log e}, \quad i = 1, 2.$$

Let us denote $\frac{\lambda_1}{1+\lambda_1} = s$ and remember that $\bar{v}_1 + \bar{v}_2 = 1$, then as functions of parameter $s \in (0, 1)$:

$$\bar{v}_1 = \frac{w_1^s}{w_1^s + w_2^s}, \quad \bar{v}_2 = \frac{w_2^s}{w_1^s + w_2^s}.$$

From the third condition in (46) we obtain the parametric expressions for E and $R_{sp}(E, W)$:

$$E(s) = \frac{w_1^s}{w_1^s + w_2^s} \log \frac{w_1^{s-1}}{w_1^s + w_2^s} + \frac{w_2^s}{w_1^s + w_2^s} \log \frac{w_2^{s-1}}{w_1^s + w_2^s},$$

$$R_{sp}(s) = 1 + \frac{w_1^s}{w_1^s + w_2^s} \log \frac{w_1^s}{w_1^s + w_2^s} + \frac{w_2^s}{w_1^s + w_2^s} \log \frac{w_2^s}{w_1^s + w_2^s}.$$

It is not complicated to see that we arrived to the same relation between R_{sp} and E as that given in Theorem 5.8.3 of the Gallager's book [13].

V. COMPARISON OF THE BOUNDS FOR E -CAPACITY

Lemma 10: For given DMC W , for type P and numbers $0 \leq E' \leq E$

$$R_r(P, E, W) = \min_{E': E' \leq E} |R_{sp}(P, E', W) + E' - E|^+.$$

Proof: Applying definitions (10) and (12) we see:

$$R_r(P, E, W) = \min_{V: D(V||W|P) \leq E} |I_{P,V}(X \wedge Y) + D(V||W|P) - E|^+ =$$

$$= \min_{E': E' \leq E, V: D(V||W|P) = E'} |I_{P,V}(X \wedge Y) + E' - E|^+ =$$

$$= \min_{E': E' \leq E} |R_{sp}(P, E', W) + E' - E|^+.$$

Lemma 11: Involving

$$E_{cr}(P, W) = \min\{E : \frac{\partial R_{sp}(P, E, W)}{\partial E} \geq -1\}.$$

we can write for all $E > 0$

$$R_r(P, E, W) = \begin{cases} R_{sp}(P, E, W), & \text{if } E \leq E_{cr}(P, W), \\ |R_{sp}(P, E_{cr}(P, W), W) + E_{cr}(P, W) - E|^+, & \text{if } E \geq E_{cr}(P, W). \end{cases}$$

Proof: Since function $R_{sp}(P, E, W)$ is convex by E then for the values of $E \leq E_{cr}(P, W)$ the slope of the tangent is not greater than -1 , and for $E \geq E_{cr}(P, W)$, it is equal or greater than -1 . In other words

$$\frac{R_{sp}(P, E, W) - R_{sp}(P, E', W)}{E - E'} \leq -1, \quad \text{when } E' < E \leq E_{cr}(P, W),$$

from where

$$R_{sp}(P, E, W) + E < R_{sp}(P, E', W) + E',$$

and consequently

$$\min_{E': E' \leq E \leq E_{cr}(P, W)} R_{sp}(P, E', W) + E' = R_{sp}(P, E, W) + E.$$

We obtain from this equality and Lemma 10 the statement of the Lemma for the case $E \leq E_{cr}(P, W)$. Now, if $E_{cr}(P, W) \leq E' < E$, then

$$\frac{R_{sp}(P, E, W) - R_{sp}(P, E', W)}{E - E'} \geq -1,$$

or

$$R_{sp}(P, E, W) + E \geq R_{sp}(P, E', W) + E',$$

and consequently

$$\min_{E': E_{cr}(P, W) \leq E'} R_{sp}(P, E', W) + E' = R_{sp}(P, E_{cr}(P, W), W) + E_{cr}(P, W).$$

Again, using Lemma 10 and the last equality we obtain that for the case $E \geq E_{cr}(P, W)$

$$\begin{aligned} R_r(P, E, W) &= \min\left\{\min_{E': E_{cr}(P, W) \leq E' < E} |R_{sp}(P, E', W) + E' - E|^+, \right. \\ &\quad \left. \min_{E': E' \leq E_{cr}(P, W)} |R_{sp}(P, E', W) + E' - E|^+ \right\} = \\ &= |R_{sp}(P, E_{cr}(P, W), W) + E_{cr}(P, W) - E|^+. \end{aligned}$$

Thus in the interval $(0, E_{cr}(P, W)]$ the functions $R(P, E, W)$ are exactly determined by

$$R(P, E, W) = R_{sp}(P, E, W) = R_r(P, E, W),$$

and in the interval $(0, E_{cr}(W))$,

$$R(E, W) = R_{sp}(E, W) = R_r(E, W).$$

So the proof of Theorem 3 is completed.

REFERENCES

- [1] E. A. Haroutunian, "Upper estimate of transmission rate for memoryless channel with countable number of output signals under given error probability exponent", (in Russian), *3th Conf. on Coding theory and Information Transmission. Uzhgorod-Tashkent.* pp. 83–86, 1967.
- [2] E. A. Haroutunian, "Combinatorial method of construction of the upper bound for E -capacity", (in Russian), *Mezhevuz. Sb. Nouch. Trudov. Matematika.* Yerevan, vol. 1. pp. 213–220, 1982.

- [3] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems" *IEEE Trans. on Inform. Theory*, vol. 27, no. 1, pp. 5–12, 1981.
- [4] R. L. Dobrushin, "Asymptotic bounds of the probability of error for the transmission of messages over a memoryless channel with a symmetric transition probability matrix", (in Russian), *Teor. Veroyatnost. i Primenen.*, vol. 7, no. 3, pp. 283–311, 1962.
- [5] R. G. Gallager, "A simple derivation of the coding theorems and some applications", *IEEE Trans. Inform. Theory*, vol. 11, no. 1, pp. 3–18, 1965.
- [6] E. A. Haroutunian, "Estimates of the error probability exponent for a semicontinuous memoryless channel", (in Russian), *Problemy Peredachi Informatcii*, vol. 4, no. 4, pp. 37–48, 1968.
- [7] R. E. Blahut, "Hypothesis testing and information theory", *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 405–417, 1974.
- [8] A. G. Dyachkov, "Bounds to average error probability for the fixed-composition code ensemble", (in Russian), *Problemy Peredachi Informatcii*, vol. 16, no. 4, pp. 3–8, 1980.
- [9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.
- [10] R. E. Blahut, *Principles and Practice of Information Theory*. Addison-Wesley, Reading, MA, 1987.
- [11] E. A. Haroutunian and H. Belbashir, "Lower bound of the optimal transmission rate depending on given error probability exponent for discrete memoryless channel and for asymmetric broadcast channel", (in Russian), *Abstracts of papers of Sixth International Symposium on Information Theory, Tashkent, USSR*, vol. 1, pp. 19-21, 1984.
- [12] E. A. Haroutunian and M. E. Haroutunian, *Information Theory*, (in Armenian), Yerevan State University, 104 p., 1987.
- [13] R. G. Gallager, *Information Theory and Reliable Communication*, New York, Wiley, 1968.